

Blundell, Richard; Horowitz, Joel L.; Pary, Matthias

**Working Paper**

## Measuring the price responsiveness of gasoline demand

cemmap working paper, No. CWP11/09

**Provided in Cooperation with:**

The Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Blundell, Richard; Horowitz, Joel L.; Pary, Matthias (2009) : Measuring the price responsiveness of gasoline demand, cemmap working paper, No. CWP11/09, Centre for Microdata Methods and Practice (cemmap), London, <https://doi.org/10.1920/wp.cem.2009.1109>

This Version is available at:

<https://hdl.handle.net/10419/64770>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Measuring the price responsiveness of gasoline demand

---

**Richard Blundell**  
**Joel L. Horowitz**  
**Matthias Parey**

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP11/09

## Measuring the Price Responsiveness of Gasoline Demand

by

Richard Blundell  
Joel L. Horowitz  
Matthias Parey\*

April 2009

### Abstract

This paper develops a new method for estimating the demand function for gasoline and the deadweight loss due to an increase in the gasoline tax. The method is also applicable to other goods. The method uses shape restrictions derived from economic theory to improve the precision of a nonparametric estimate of the demand function. Using data from the U.S. National Household Travel Survey, we show that the restrictions are consistent with the data on gasoline demand and remove the anomalous behavior of a standard nonparametric estimator. Our approach provides new insights about the price responsiveness of gasoline demand and the way responses vary across the income distribution. We reject constant elasticity models and find that price responses vary non-monotonically with income. In particular, we find that low- and high-income consumers are less responsive to changes in gasoline prices than are middle-income consumers.

JEL: D120, H310, C140

\* Blundell: Department of Economics, University College London, and Institute for Fiscal Studies, Gower Street, London WC1E 6BT, UK (email: r.blundell@ucl.ac.uk); Horowitz: Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208, USA (email: joel-horowitz@northwestern.edu); Parey: Department of Economics, University College London, and Institute for Fiscal Studies, Gower Street, London WC1E 6BT, UK (email: m.parey@ucl.ac.uk). We thank seminar participants at IFS and UCL for helpful comments. The paper is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS. Financial support from the ESRC is gratefully acknowledged. The research of Joel L. Horowitz was supported in part by NSF grants SES 0352675 and SES-0817552.

## 1. Introduction

This paper describes a new method for estimating a demand function for gasoline and the welfare costs of changes in gasoline prices. The method is also applicable to other goods. In the U.S., as in many other countries, the price of gasoline rose rapidly from 1998 until mid 2008. Figure 1 shows the how the average price of gasoline in the U.S. has varied over the last three decades. Prices began rising steeply in about 1998 following a period of price stability that began in about 1986. Between March 2007 and March 2008, the average gasoline price increased by 25.7 percent in nominal terms.<sup>1</sup> In real terms, gasoline prices reached levels similar to those seen during the second oil crisis of 1979-1981. Although prices have decreased since mid 2008, due at least in part to the global economic downturn, many observers expect prices to rise again in the future as economic activity increases.

The measurement of the welfare consequences of price changes begins with estimating the demand function for the good in question. This is often done by using a linear model in which the dependent variable is the log of demand and the explanatory variables are the logs of price and income. This model is easy to interpret because it gives constant income and price elasticities, but it is rejected by our data. Table 1 presents the results of estimating a constant elasticity model of gasoline demand for a class of households in the U.S. The data are from the National Household Travel Survey (NHTS). We describe the NHTS further in Section 3. RESET specification tests reject the constant-elasticity model. Further analysis that is described in Section 4 reveals that adding an interaction term to the constant elasticity model does not correct the specification error. This motivates us to use nonparametric estimation methods. Hausman and Newey (1995) also used nonparametric methods to estimate gasoline demand.

Deviations from the constant-elasticity model are not simply a technical concern. It is likely to matter greatly how peoples' responses to prices vary according to the price level and over the income distribution. Therefore, a flexible modeling approach such as nonparametric regression seems attractive. However, nonparametric regression can yield implausible and erratic estimates. Figure 2 shows nonparametric estimates of gasoline demand as a function of price at three points across the income distribution. The estimates are obtained from the NHTS data. Details of the estimation method are presented in Section 2 of this paper. The figure gives some overall indication of downward sloping demand curves with slopes that differ across the income distribution but there are parts of the estimated demand curves that are upward sloping and, therefore, implausible. We interpret the implausible shapes of the curves in Figure 2 as indicating

---

<sup>1</sup> Own calculation based on EIA (2008b, Table 9.4).

that fully nonparametric methods are too imprecise to provide useful estimates of gasoline demand functions with our data.

One way of dealing with this problem is to impose a parametric form such as log-log linearity on the demand function. But any parametric form is essentially arbitrary and, as will be discussed further in Section 4, may be misspecified in ways that produce seriously erroneous results. As a compromise between the desire for flexibility and the need for structure, one may use a semiparametric model, such as a partially-linear or single-index model. These impose parametric restrictions on some aspects of the function of interest but leave other parts unrestricted. In this paper, we take a different approach and impose structure through shape restrictions based on economic theory. Specifically, we impose the Slutsky restriction of consumer theory on an otherwise fully nonparametric estimate of the demand function. We show that this approach yields well-behaved estimates of the demand function and price responsiveness across the income distribution while avoiding the use of arbitrary and possibly misspecified parametric or semiparametric models. We implement our approach by making use of a kernel-type estimator in which observations are weighted in a way that ensures satisfaction of the Slutsky restrictions. This maintains the flexibility of nonparametric regression while using restrictions of economic theory to avoid implausible estimation results. The constrained nonparametric estimates are consistent with observed behavior and provide intuitively plausible, well-behaved descriptions of price responsiveness across the income distribution.

One important use of demand function estimates is to compute deadweight loss (DWL) measures of tax policy interventions. For some interventions, we show that reliance on the unrestricted nonparametric estimate results in DWL estimates that have incorrect signs and are, therefore, nonsensical. Our constrained estimator deals with this problem in a way that is consistent with economic theory.

We find that there is substantial variation in price sensitivity across both price and income. In particular, we find that price responses are non-monotonic in income. Our estimates indicate that households at the median of the income distribution respond more strongly to an increase in prices than do households at the lower or upper income group. We do not speculate on why this is the case, but we show that it implies that our DWL measure is typically higher at the median of the income distribution than in the lower or upper income group.

Section 2 explains our approach to nonparametric estimation of demand functions and DWL subject to the Slutsky shape restrictions. Section 3 describes the NHTS data. Section 4 presents the estimates of the demand function and shows how price responsiveness varies across

the income distribution. Section 4 also presents the DWLs associated with several price changes and shows how they vary across the income distribution. Section 5 concludes.

## 2. Shape Restrictions and the Estimation of Demand and Deadweight Loss

We begin this section by describing our approach to estimating the demand function subject to the Slutsky shape restriction. Then we describe how we estimate the DWL of a tax-induced price increase.

The Slutsky condition is an inequality constraint on the demand function. Our method for estimating the demand function nonparametrically subject to this constraint is adapted from Hall and Huang (2001), who present a nonparametric kernel estimator of a conditional mean function subject to a monotonicity constraint. We replace their monotonicity constraint with the Slutsky condition. To describe our estimator, let  $Q$ ,  $P$ , and  $Y$ , respectively, denote the quantity of gasoline demanded by an individual, the price paid, and the individual's income. We assume that these variables are related by

$$(1) \quad Q = g(P, Y) + U,$$

where  $g$  is a function that satisfies smoothness conditions and the Slutsky restriction but is otherwise unknown, and  $U$  is an unobserved random variable satisfying  $E(U | P = p, Y = y) = 0$  for all  $p$  and  $y$ . Our aim is to estimate  $g(p, y)$  nonparametrically subject to the Slutsky constraint

$$(2) \quad \frac{\partial g(p, y)}{\partial p} + g(p, y) \frac{\partial g(p, y)}{\partial y} \leq 0.$$

The data are observations  $\{Q_i, P_i, Y_i : i = 1, \dots, n\}$  for  $n$  randomly sampled individuals. A fully nonparametric estimate of  $g$  that does not impose the Slutsky restriction can be obtained by using the Nadaraya-Watson kernel estimator (Nadaraya 1964, Watson 1964). The properties of this estimator are summarized in Härdle (1990). We call it the unconstrained nonparametric estimator, denoted by  $\hat{g}_U$ , because it is not constrained by (2). The estimator is

$$(3) \quad \hat{g}_U(p, y) = \frac{1}{nh_p h_y \hat{f}(p, y)} \sum_{i=1}^n Q_i K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right),$$

where

$$\hat{f}(p, y) = \frac{1}{nh_p h_y} \sum_{i=1}^n K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right),$$

$K$  is a bounded, differentiable probability density function that is supported on  $[-1,1]$  and is symmetrical about 0, and  $h_p$  and  $h_y$  are bandwidth parameters.

Owing to the effects of random sampling errors,  $\hat{g}_U$  does not necessarily satisfy (2) even if  $g$  does satisfy this condition. Following Hall and Huang (2001), we solve this problem by replacing  $\hat{g}_U$  with the weighted estimator

$$(4) \quad \hat{g}_C(p, y) = \frac{1}{h_p h_y \hat{f}(p, y)} \sum_{i=1}^n w_i Q_i K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right),$$

where  $\{w_i : i = 1, \dots, n\}$  are non-negative weights satisfying  $\sum_{i=1}^n w_i = 1$  and the subscript  $C$  indicates that the estimator is constrained by the Slutsky condition. The weights are obtained by solving the optimization problem

$$(5) \quad \underset{w_1, \dots, w_n}{\text{minimize}} : D(w_1, \dots, w_n)$$

subject to

$$\frac{\partial \hat{g}_C(p_j, y_j)}{\partial p} + \hat{g}_C(p, y) \frac{\partial \hat{g}_C(p_j, y_j)}{\partial y} \leq 0; \quad j = 1, \dots, J; \quad ,$$

$$\sum_{i=1}^n w_i = 1,$$

and

$$w_i \geq 0; \quad i = 1, \dots, n,$$

where  $\{p_j, y_j : j = 1, \dots, J\}$  is a grid of points in the  $(p, y)$  plane. The objective function is the following measure of the “distance” of the weights from the values  $w_i = 1/n$  corresponding to the Nadaraya-Watson estimator:

$$D(w_1, \dots, w_n) = n - \sum_{i=1}^n (nw_i)^{1/2}.$$

When  $w_i = 1/n$  for all  $i = 1, \dots, n$ ,  $\hat{g}_C(p_j, y_j) = \hat{g}_U(p_j, y_j)$  for all  $j = 1, \dots, J$ . Thus, the weights minimize the distance of the constrained estimator from the unconstrained one. The constraint is not binding at points  $(p_j, y_j)$  that satisfy (2). In the empirical application described in Section 4, we solve (5) by using the nonlinear programming algorithm E04UCF from the NAG Library. The bandwidths are selected using a method that is described in Section 4. In some applications,

it may be desirable to impose the restriction that the good in question is normal. This can be done by adding the constraints  $\partial \hat{g}_C(p_j, y_j) / \partial y \geq 0$  to (5), but we do not take this step here.

We now describe our method for estimating the DWL of a tax. Let  $E(p)$  denote the expenditure function at price  $p$  and some reference utility level. The DWL of a tax that changes the price from  $p^0$  to  $p^1$  is

$$(6) \quad L(p^0, p^1) = E(p^1) - E(p^0) - (p^1 - p^0)g[p^1, E(p^1)].$$

We estimate this by

$$(7) \quad \hat{L}(p^0, p^1) = \hat{E}(p^1) - \hat{E}(p^0) - (p^1 - p^0)\hat{g}[p^1, \hat{E}(p^1)],$$

where  $\hat{E}$  is an estimator of the expenditure function and  $\hat{g}$  may be either  $\hat{g}_U$  or  $\hat{g}_C$ . We obtain  $\hat{E}$  by solving the differential equation

$$(8) \quad \frac{d\hat{E}(t)}{dt} = \hat{g}[p(t), \hat{E}(t)] \frac{dp(t)}{dt},$$

where  $[p(t), \hat{E}(t)]$  ( $0 \leq t \leq 1$ ) is a price-(estimated) expenditure path. We solve this equation along a grid of points by using Euler's method (Ascher and Petzold 1998). We have found this method to be quite accurate in numerical experiments.

Inference with the constrained estimator  $\hat{g}_C$  is difficult because the estimator's asymptotic distribution is very complicated in regions where (2) is a binding constraint (strict equality). However, if we assume that (2) is a strict inequality in the population, then violation of the Slutsky condition by  $\hat{g}_U$  is a finite-sample phenomenon, and we can use  $\hat{g}_U$  to carry out asymptotically valid inference. We use the bootstrap to obtain asymptotic joint confidence intervals for  $g(p, y)$  on a grid of  $(p, y)$  points and to obtain confidence intervals for  $L$ . The bootstrap procedure is as follows.

1. Generate a bootstrap sample  $\{Q_i^*, P_i^*, Y_i^* : i = 1, \dots, n\}$  by sampling the data randomly with replacement.

2. Use this sample to estimate  $g(p, y)$  on a grid of  $(p, y)$  points without imposing the Slutsky constraint. Also, estimate  $L$ . Denote the bootstrap estimates by  $\hat{g}_U^*$  and  $L^*$ .

3. Form percentile confidence intervals for  $L$  by repeating steps 1-2 many times. Also, use the bootstrap samples to form joint percentile- $t$  confidence intervals for  $g$  on the grid of points  $\{p_j, y_j : j = 1, \dots, J\}$ . The joint confidence intervals at a level of at least  $1 - \alpha$  are

$$(9) \quad \hat{g}_U(p_j, y_j) - z_\alpha(p_j, y_j)\hat{\sigma}(p_j, y_j) \leq g(p_j, y_j) \leq \hat{g}_U(p_j, y_j) + z_\alpha(p_j, y_j)\hat{\sigma}(p_j, y_j),$$



where

$$(10) \quad \hat{\sigma}^2(p, y) = \frac{B_K}{[nh_p h_y \hat{f}(p, y)]^2} \sum_{i=1}^n \hat{U}_i^2 K\left(\frac{p - P_i}{h_p}\right) K\left(\frac{y - Y_i}{h_y}\right),$$

$$\text{with } B_K = \int K(v)^2 dv \text{ and } \hat{U}_i = Q_i - \hat{g}_U(P_i, Y_i),$$

is a consistent estimate of  $\text{Var}[\hat{g}_U(p, y)]$ . The coefficient  $z_\alpha(p_j, y_j)$  is chosen following the approach in Härdle and Marron (1991) for computing joint confidence intervals. For this purpose, we partition the grid into intervals of  $2h_p$ . Within each of these  $M$  neighborhoods,  $z_\alpha(p_j, y_j)$  is the solution to

$$P^* \left[ \frac{|\hat{g}_U^*(p_j, y_j) - \hat{g}_U(p_j, y_j)|}{\hat{\sigma}^*(p_j, y_j)} \leq z_\alpha(p_j, y_j) \right] = 1 - \beta,$$

where  $P^*$  is the probability measure induced by bootstrap sampling, and  $\hat{\sigma}^*(p, y)$  is the version of  $\hat{\sigma}(p, y)$  that is obtained by replacing  $\hat{U}_i$ ,  $P_i$ , and  $Y_i$  in (10) by their bootstrap analogs, and  $\beta$  is a parameter. We then choose  $\beta$  such that the simultaneous size in each neighborhood equals  $1 - \frac{\alpha}{M}$ . As Härdle and Marron (1991) show using the Bonferroni inequality, the resulting intervals over the full grid form simultaneous confidence intervals at a level of at least  $1 - \alpha$ . Hall (1992) shows that the bootstrap consistently estimates the asymptotic distribution of the Studentized form of  $\hat{g}_U$ . It is necessary to undersmooth  $\hat{g}_U$  and  $\hat{g}_U^*$  (that is, use smaller than asymptotically optimal bandwidths) in (9) and step 2 of the bootstrap procedure to obtain a confidence interval that is centered at  $g$ . We discuss bandwidth selection in Section 4.

### 3. Data

Our analysis is based on the 2001 National Household Travel Survey. The NHTS was sponsored by the Bureau of Transportation Statistics and the Federal Highway Administration. The data were collected through a telephone survey of the civilian, non-institutionalized population of the U.S. The survey was conducted between March 2001 and May 2002 (ORNL 2004, Ch. 3). The telephone interviews were complemented with written travel diaries and odometer readings.

The variables used in our study are annual gasoline consumption, the gasoline price, and household income. Gasoline consumption is derived from odometer readings and estimates of the fuel efficiencies of vehicles. Details of the computations are described in ORNL (2004,

Appendices J and K). The gasoline price for a given household is the average price in dollars per gallon, including taxes, in the county where the household is located. This price variable is a county average, rather than the price actually paid by a household. It precludes an intra-county analysis (see Schmalensee and Stoker 1999) but does capture variation in prices consumers face in different regions.

Household income in dollars is available in 18 groups. In our analysis, we assign each household an income equal to the midpoint of its group. The highest group, consisting of incomes above \$100,000, is assigned an income of \$120,000.<sup>2</sup> To investigate how price responsiveness of gasoline demand varies across the income distribution, we focus on three income levels of interest: a middle income group at \$57,500, which corresponds to median income in our sample, a low income group (\$42,500), which corresponds to the first quartile and a high income group (\$72,500)<sup>3</sup>. To obtain gasoline demand at the household level, we aggregate vehicle gasoline expenditure in dollars and gasoline consumption in gallons over multi-car households. We divide the household gasoline expenditure by the quantity of gasoline consumed to obtain the household's gasoline price. We do not investigate the errors-in-variables issues raised by the use of county-average prices or the interval censoring issues raised by the grouping of household incomes in the data. These potentially important issues are left for future research.

We exclude from our analysis households where the number of drivers is zero or whose income, gasoline cost, or annual gasoline consumption is not reported. We also exclude households that are located in Hawaii. In addition, we restrict our sample to households with a white respondent, two or more adults, and at least one child under 16 years of age. We take vehicle ownership as given and do not investigate how changes in prices affect vehicle purchases or how vehicle ownership varies across the income distribution (Poterba 1991; West 2004; Bento, Goulder, Henry, Jacobsen, and von Haefen 2005; Bento, Goulder, Jacobsen, and von Haefen 2009). The results of Bento, et al. (2005) indicate that over 95 percent of the reduction in gasoline demand in response to price changes is due to changes in miles traveled rather than fleet composition. We limit attention to vehicles that use gasoline as fuel, rather than diesel, natural

---

<sup>2</sup> Assuming log-normality of income, we have estimated the corresponding mean and variance by using a simple tobit model, right-censored at \$100,000. Excluding households with very high incomes above \$150,000, the median income in the upper group corresponds to about \$120,000.

<sup>3</sup> The income point \$72,500 occupies the 59.6-63.3th percentile. This point was chosen to avoid the problems created by the interval nature of the income variable which becomes especially important in the upper quartile of the income distribution: income brackets are relatively narrow (with widths of \$5,000) up to \$80,000, but substantially wider for higher incomes. However, estimates using higher quantiles yielded similar results and did not change our conclusions on price responsiveness across the income distribution.

gas, or electricity. The resulting sample consists of 5,257 observations. Table 2 shows summary statistics.

#### 4. Estimates of Demand Responses

##### a. The constant elasticity model

We begin by using ordinary least squares to estimate the following log-log linear demand model:

$$(11) \quad \log Q = \beta_0 + \beta_1 \log P + \beta_2 \log Y + U; \quad E(U | P = p, Y = y) = 0.$$

This constant elasticity model is one of the most frequently estimated (e.g., Dahl 1979; Hughes, Knittel, and Sperling 2008). It has been criticized on many grounds (e.g., Deaton and Muellbauer 1980) but its simplicity and frequent use make it a useful parametric reference model. Later in this section, we compare the estimates obtained from model (11) with those obtained from the nonparametric analysis.

The estimates of the coefficients of (11) are shown in Table 1. They imply a price-elasticity of demand of -0.88 and an income elasticity of 0.29. These estimates are similar to those reported by others. Hausman and Newey (1995) report estimates of -0.81 and 0.37, respectively, for price and income elasticities based on U.S. data collected between 1979 and 1988. Schmalensee and Stoker (1999) report price elasticities between -0.72 and -1.13 and income elasticities between 0.12 and 0.33, depending on the survey year and control variables, in their specifications without regional fixed effects. Yatchew and No (2001) estimate an income elasticity of 0.27 using Canadian data for 1994-1996 and a model that does not include the price of gasoline. West (2004) reports a mean price elasticity of -0.89 using 1997 data.

Although the estimates we obtain from model (11) are similar to those reported by others, there is evidence that (11) is misspecified. We tested (11) for misspecification with Ramsey's (1969) RESET test. This test consists of adding powers of the predicted values of  $\log Q$  to the model, re-estimating the resulting augmented model, and testing the hypothesis that the coefficients of the additional regressors are zero. Rejection of this hypothesis indicates that the original model is misspecified. We carried out this test twice, once with the squares and cubes of the predicted  $\log Q$  values added to the model (RESET3 in Table 1) and once with the squares, cubes, and fourth powers of the  $\log Q$ 's added (RESET4). As can be seen from Table 1, both versions of RESET reject model (11) at the 0.05 level. Thus, we conclude that model (11) is misspecified.

West (2004) found evidence for dependence of the price elasticity on income. Accordingly, we added the interaction term  $(\log P)(\log Y)$  to model (11). The resulting augmented model is also rejected at the 0.05 level by the RESET tests. Conceivably adding further powers and interactions of  $\log P$  and  $\log Y$  would yield a model that is not rejected by RESET. However, this kind of informal specification search leads to inconsistent estimators whose properties are unknown. Nonparametric estimators, by contrast, are consistent.

b. Unconstrained nonparametric estimates

Our unconstrained nonparametric estimates of the demand function,  $\hat{g}_U$ , are displayed in Figure 2. They were obtained by using the Nadaraya-Watson kernel estimator with a biweight kernel. In principle, the bandwidths  $h_p$  and  $h_y$  can be chosen by applying least-squares cross-validation (Härdle 1990) to the entire data set, but this yields bandwidths that are strongly influenced by low-density regions. To avoid this problem, we used the following method to choose  $h_p$  and  $h_y$ . We are interested in  $g(p, y)$  for  $y$  values corresponding to our three income groups and price levels between the 5th and 95th percentiles of the observed prices. We defined three price-income rectangles consisting of prices between the 5th and 95th percentiles and incomes within 0.5 of each income level of interest (measured in logs). We then applied least-squares cross-validation to each price-income rectangle separately to obtain bandwidth estimates appropriate to each rectangle. This procedure yielded  $(h_p, h_y) = (0.0431, 0.2143)$  for the lower income group,  $(0.0431, 0.2061)$  for the middle income group, and  $(0.0210, 0.2878)$  for the upper income group. The estimation results are not sensitive to modest variations in the dimensions of the price-income rectangles. As was discussed in Section 2,  $\hat{g}_U$  and  $\hat{g}_U^*$  must be undersmoothed to obtain properly centered confidence intervals. To this end we multiplied each of the foregoing bandwidths by 0.8 when computing confidence intervals.

Figure 2 shows several instances in which the nonparametric estimate of the (Marshallian) demand function is upward sloping. This anomaly is also present in the results of Hausman and Newey (1995). The theory of the consumer requires the compensated demand function to be downward sloping. Combined with a positive income derivative, an upward-sloping Marshallian demand function implies an upward-sloping compensated demand function and, therefore, is inconsistent with the theory of the consumer. At the median income, our nonparametric estimate of  $\partial g / \partial y$  is positive over the range of prices of interest except for the two lowest grid points. Therefore, the nonparametric estimates are inconsistent with consumer theory. As is discussed in more detail in Section 4c, we believe this result to be an artifact of

random sampling errors and the consequent imprecision of the unconstrained nonparametric estimates. This motivates the use of the constrained estimation procedure, which increases estimation precision by imposing the Slutsky condition.

c. Nonparametric estimates under the Slutsky condition

Figure 3 shows the nonparametric estimates of the demand function,  $\hat{g}_C$ , at each of the three income levels of interest (solid dots). These estimates are constrained to satisfy the Slutsky condition and were obtained using the methods described in Section 2. For comparison, the figure also shows the unconstrained nonparametric estimates,  $\hat{g}_U$  (open dots). The solid lines in Figure 3 connect the endpoints of joint 90% confidence intervals for  $g(p, y)$ . These were obtained using the bootstrap procedure described in Section 2.

In contrast to the unconstrained estimates, the constrained estimates are downward sloping everywhere. The constrained estimates are also less wiggly than the unconstrained ones. In contrast to ad hoc “ironing procedures” for producing monotonic estimates,  $\hat{g}_C$  is consistent with the theory of the consumer and everywhere differentiable. This is important for estimation of DWL. The 90% confidence bands shown in Figure 3 contain both the constrained and unconstrained estimates. This is consistent with our view that the anomalous behavior of the unconstrained estimates is due to imprecision of the unconstrained estimator. It also indicates that the Slutsky constraint is consistent with the data.

The results in Figure 3 indicate that the middle income group is more sensitive to price changes than are the other two groups. In particular, the slope of the constrained estimate of  $g$  is noticeably larger for the middle group than for the other groups. This, in turn, suggests that the DWL of a tax increase is larger for the middle income group than for the others. We investigate this further in Section 4d.

d. Estimates of deadweight loss

We now investigate the DWLs associated with several increases in gasoline taxes. The increases considered in the literature typically are quite large and often out of the support of the data. We start with an intervention that moves prices from the 5th to the 95th percentile of the price distribution in our sample. Historically observed tax changes in the U.S. tend to be much smaller than this, possibly due to the political difficulty of implementing large tax increases. To reflect the kind of intervention a legislature might actually consider, we also look at smaller interventions in which the price increases by \$0.05. As is well known, DWL increases with the square of the tax rate (e.g., Auerbach 1985), so the DWL estimates are very different for the two types of interventions.

We compute DWL as follows. Over the range of the intervention, we evaluate the Marshallian demand estimates presented in the previous section for the three estimators (parametric, unconstrained nonparametric, and constrained nonparametric) on a grid of 61 points. We then use this demand estimate and the corresponding derivatives to compute the expenditure function and DWL by following the methods described in Section 2.

We study DWL relative to tax paid, which we interpret as a “price” for raising tax revenue. We refer to this measure as relative DWL. Results are shown in Table 3.<sup>4</sup> Each panel of the table corresponds to one intervention. Intervention I moves prices from the 5th to the 95th price percentile in our data. The differences in the demand estimates between the different estimation methods translate into differences in relative DWLs. Comparing across income levels, the log-log linear model estimates relative DWL to be almost identical for the three income groups and indicates that the cost of taxation is about 7.6% of revenue raised for intervention I, irrespective of income level. In contrast, the constrained nonparametric estimates indicate that the cost of taxation is higher for the middle income group than for the other two groups. This result is consistent with our earlier finding that the middle income group is more responsive to price changes than are the other groups. The result also illustrates how the functional form assumptions of the parametric model affect estimates of consumer behavior and the effects of taxation.

We also estimate the DWLs associated with taxes that increase the price by \$0.05 from several different initial values. Intervention II increases the price from \$1.22 to \$1.27, Intervention III from \$1.27 to \$1.32, and Intervention IV from \$1.32 to \$1.37. The results are shown in Table 3. The DWLs obtained from the log-log linear parametric model of the demand function are virtually constant across incomes. The DWLs obtained from the unconstrained nonparametric estimate of demand are sometimes negative. This anomalous result occurs because, due to random sampling errors, the unconstrained estimate of the demand function does not decrease monotonically and does not satisfy the integrability conditions of consumer theory. The constrained nonparametric model yields DWL estimates that are positive and, in some cases, more than double those obtained from the parametric model.

One can also study DWL relative to income so as to reflect the household's utility loss relative to available resources. The results for this analysis are shown in Table 4. The estimates from the parametric model and constrained nonparametric model give different indications of the effects of the tax increase across income groups. The parametric estimates indicate that the relative utility loss increases as income decreases. However, the constrained nonparametric

---

<sup>4</sup> Confidence intervals for the unconstrained and the parametric model are reported in Table 5.

estimates indicate that the relative utility loss is greater for the middle income group than for the other groups.

## 5. Conclusions

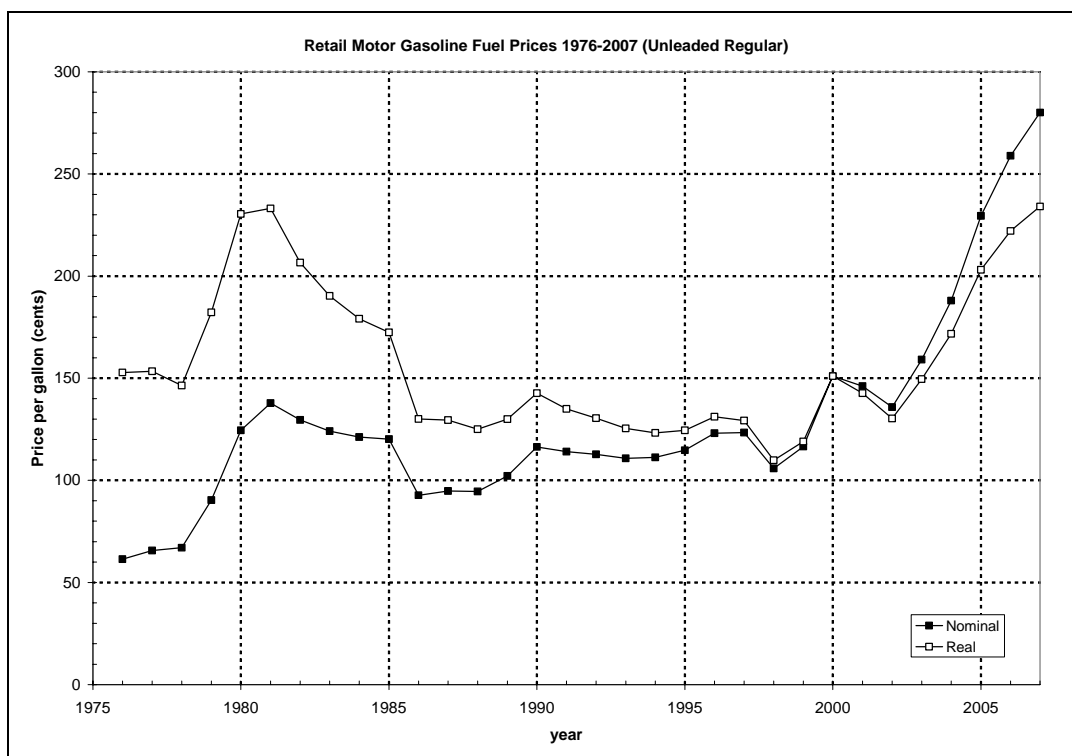
Simple parametric models of demand functions can yield misleading estimates of price sensitivity and welfare measures such as DWL, owing to misspecification. Fully nonparametric estimation of demand reduces the risk of misspecification but, because of the effects of random sampling errors, can yield imprecise estimates with anomalous properties such as non-monotonicity. This paper has shown that these problems can be overcome by constraining nonparametric estimates to satisfy the Slutsky condition of economic theory. This stabilizes the nonparametric estimates without the need for parametric or other restrictions that have no basis in economic theory.

We have implemented this approach by using a modified kernel estimator that weights the observations so as to satisfy the Slutsky restriction. To illustrate the method, we have estimated a gasoline demand function for a class of households in the U.S. We find that some simple parametric specifications are rejected by a specification test, whereas a fully nonparametric estimate of the demand function is non-monotonic. In contrast, the estimate that is constrained to satisfy the Slutsky condition is well-behaved. Moreover, the constrained nonparametric estimates show patterns of price sensitivity that are very different from those of the simple parametric model. We find price responses vary non-monotonically with income. In particular, we find that low- and high-income consumers are less responsive to changes in gasoline prices than are middle-income consumers.

We have also computed the DWLs of several increases in the price of gasoline. We find that the unconstrained nonparametric estimates sometimes yield negative DWLs, which are inconsistent with economic theory and presumably caused by imprecision of the unconstrained estimates. The constrained nonparametric estimates of DWL are positive and, in many cases, quite different from those obtained with the parametric model. Mirroring the results on price responsiveness, the DWL estimates are highest for middle income groups. These results illustrate the usefulness of nonparametrically estimating demand functions subject to the Slutsky condition.

## FIGURES

**Figure 1: Retail Motor Gasoline Price 1976-2007 (Unleaded Regular)**



Source: EIA (2008a, Table 5.24). Note: U.S. city average gasoline prices. Real values are in chained (2000) dollars based on GDP implicit price deflators. See source for details.



**Table 1: OLS regression and specification test**

dependent variable: log gasoline demand	
log price	-0.885 [0.157]**
log income	0.292 [0.015]**
constant	4.226 [0.166]**
Observations	5,257
RESET3: F-stat	5.522
RESET3: p-value	0.004**
RESET4: F-stat	4.034
RESET4: p-value	0.007**

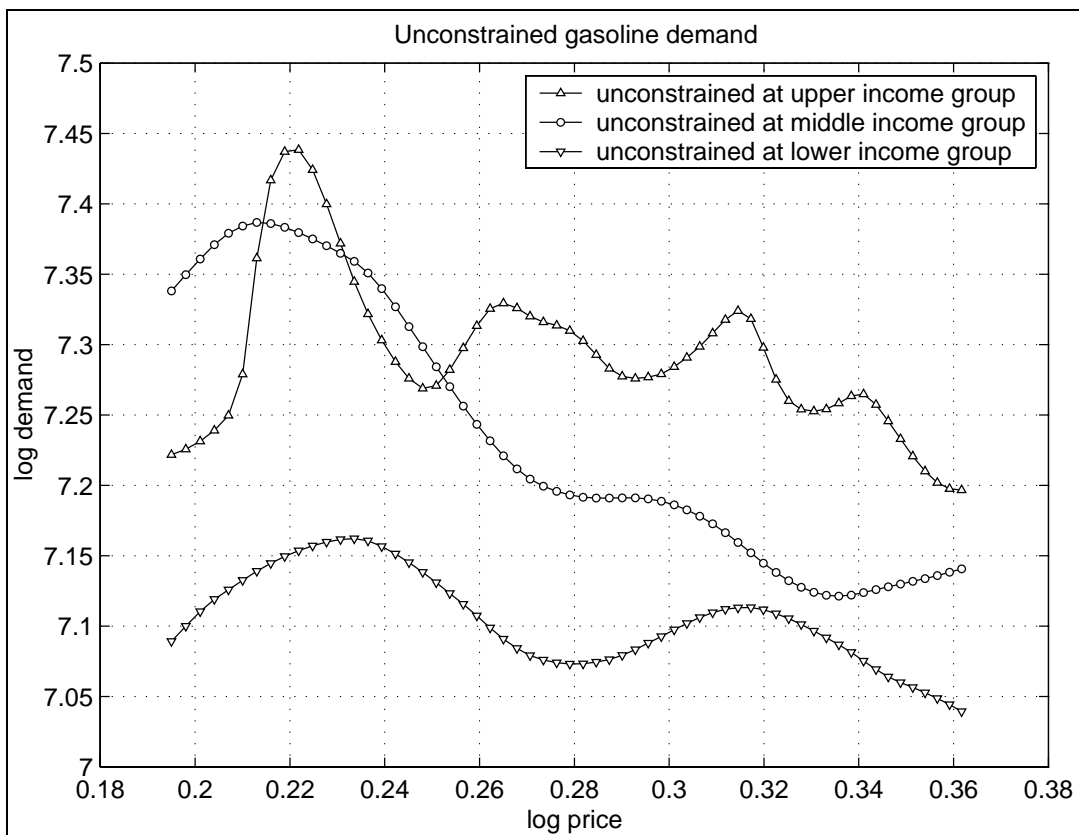
Note: Dependent variable is log of annual household gasoline demand in gallons. \* indicates significance at 5%, \*\* indicates significance at 1% level. The bottom panel reports results from the F-test of two Ramsey RESET specification tests. RESET3 refers to including second and third polynomials of the predicted values of the dependent variable, and RESET4 refers to including second to fourth polynomials. See text for details.

**Table 2: Sample descriptives**

log gasoline demand	7.168 [0.679]
log price	0.287 [0.057]
log income	10.954 [0.613]
Observations	5,257

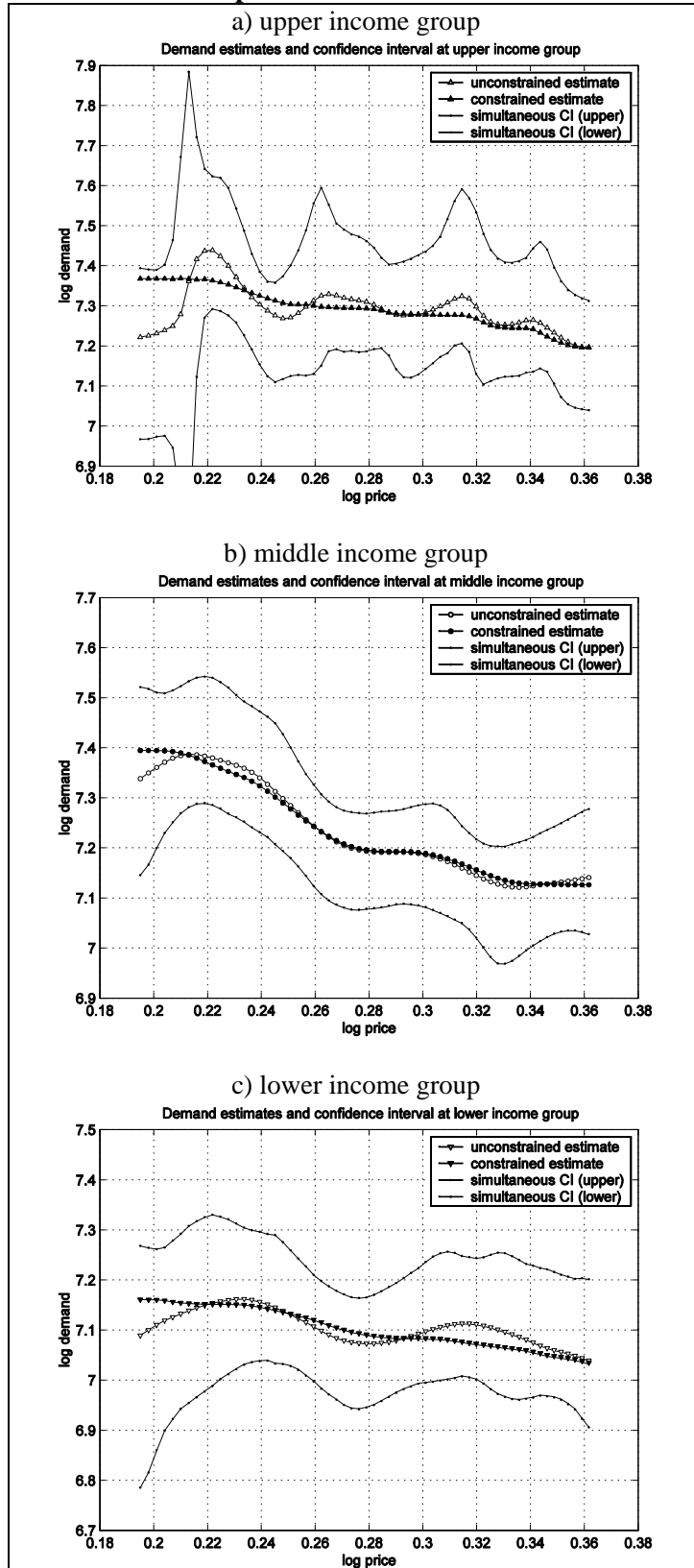
Note: Table shows means and standard deviations.

**Figure 2: Unconstrained nonparametric demand estimates**



Note: Income groups correspond to \$72,500, \$57,500, and \$42,500.

**Figure 3: Demand estimates and simultaneous confidence intervals at different points in the income distribution**



Note: Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown refer to bootstrapped symmetrical, studentized simultaneous confidence intervals with a confidence level of 10%, based on 10,000 replications. See text for details.

**Table 3: Relative Deadweight Loss estimates**

	Income	DWL (as % of tax paid)		
		unconstrained (1)	constrained (2)	log-log (3)
Intervention I (\$1.215 - \$1.436)	\$72,500	10.09 %	10.18 %	7.59 %
	\$57,500	10.09 %	11.92 %	7.58 %
	\$42,500	6.40 %	6.70 %	7.56 %
Intervention II (\$1.22 - \$1.27)	\$72,500	4.20 %	3.27 %	1.80 %
	\$57,500	3.08 %	4.50 %	1.80 %
	\$42,500	-1.33 %	0.72 %	1.79 %
Intervention III (\$1.27 - \$1.32)	\$72,500	-1.06 %	0.84 %	1.73 %
	\$57,500	6.42 %	5.74 %	1.73 %
	\$42,500	3.86 %	2.82 %	1.72 %
Intervention IV (\$1.32 - \$1.37)	\$72,500	-3.02 %	0.49 %	1.67 %
	\$57,500	2.61 %	2.07 %	1.66 %
	\$42,500	-2.23 %	0.77 %	1.66 %

Note: For each intervention, the price change considered is indicated in round brackets (in U.S. dollars). Intervention I corresponds to moving prices from the 5th to the 95th percentile in the data. Interventions II, III and IV each increase price by five U.S. cents. Deadweight Loss is shown as percentage of tax paid after the (compensated) intervention. See text for details.

**Table 4: Deadweight Loss estimates relative to household income**

	Income	DWL (relative to income) * 10 <sup>4</sup>		
		unconstrained (1)	constrained (2)	log-log (3)
Intervention I (\$ 1.215-1.436)	\$72,500	4.11	4.14	3.01
	\$57,500	4.89	5.69	3.54
	\$42,500	3.80	3.97	4.37
Intervention II (\$ 1.22-1.27)	\$72,500	0.43	0.34	0.18
	\$57,500	0.41	0.59	0.21
	\$42,500	-0.20	0.11	0.26
Intervention III (\$ 1.27-1.32)	\$72,500	-0.11	0.09	0.17
	\$57,500	0.74	0.67	0.20
	\$42,500	0.54	0.40	0.24
Intervention IV (\$ 1.32-1.37)	\$72,500	-0.32	0.05	0.16
	\$57,500	0.29	0.23	0.18
	\$42,500	-0.32	0.11	0.23

Note: For each intervention, the price change considered is indicated in round brackets (in U.S. dollars). Intervention I corresponds to moving prices from the 5th to the 95th percentile in the data. Interventions II, III and IV each increase price by five U.S. cents. Deadweight Loss is shown relative to baseline income. See text for details.

**Table 5: Confidence intervals for DWL measures**

	Income	DWL (as % of tax paid)				DWL (relative to income) * 10 <sup>4</sup>			
		unconstrained		log-log		unconstrained		log-log	
		lower	upper	lower	upper	lower	upper	lower	upper
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intervention I (\$ 1.215-1.436)	\$ 72,500	1.24 %	23.03 %	4.81 %	10.33 %	1.36	9.09	1.99	4.04
	\$ 57,500	-1.50 %	17.69 %	4.82 %	10.17 %	0.02	8.69	2.35	4.70
	\$ 42,500	-7.24 %	15.21 %	4.81 %	10.28 %	-3.50	9.54	2.90	5.88
Intervention II (\$ 1.22-1.27)	\$ 72,500	-5.34 %	16.39 %	1.17 %	2.43 %	-0.50	1.63	0.11	0.25
	\$ 57,500	-3.07 %	9.90 %	1.17 %	2.40 %	-0.37	1.36	0.13	0.28
	\$ 42,500	-10.31 %	3.18 %	1.17 %	2.42 %	-1.56	0.51	0.17	0.36
Intervention III (\$ 1.27-1.32)	\$ 72,500	-11.26 %	6.37 %	1.13 %	2.34 %	-1.14	0.74	0.11	0.23
	\$ 57,500	0.52 %	14.40 %	1.13 %	2.30 %	0.11	1.63	0.13	0.26
	\$ 42,500	-1.99 %	12.78 %	1.13 %	2.33 %	-0.19	1.75	0.16	0.33
Intervention IV (\$ 1.32-1.37)	\$ 72,500	-18.92 %	0.87 %	1.09 %	2.25 %	-2.11	0.32	0.10	0.21
	\$ 57,500	-1.81 %	7.77 %	1.09 %	2.22 %	-0.17	0.86	0.12	0.24
	\$ 42,500	-8.29 %	2.45 %	1.09 %	2.24 %	-1.20	0.41	0.15	0.30

Note: For each intervention, the price change considered is indicated in round brackets (in U.S. dollars). Intervention I corresponds to moving prices from the 5th to the 95th percentile in the data. Interventions II, III and IV each increase price by five U.S. cents. Table shows confidence intervals corresponding to estimates reported in Tables 3 and 4. Confidence intervals are computed with an undersmoothed bandwidth, based on 5,000 replications. See text for details.

## REFERENCES

- Ascher, U.M. and Petzold, L.R. (1998). *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM.
- Auerbach, A.J. (1985). The Theory of Excess Burden and Optimal Taxation, in *Handbook of Public Economics*, ed. by Auerbach, A.J., and Feldstein, M., chap. 2, pp. 61-127. Elsevier.
- Bento, A.M., Goulder, L.H., Henry, E., Jacobsen, M.R. and von Haefen, R.H. (2005). Distributional and Efficiency Impacts of Gasoline Taxes: An Econometrically-Based Multi-Market Study, *The American Economic Review*, 95(2), 282-287, Papers and Proceedings.
- Bento, A.M., Goulder, L.H., Jacobsen, M.R. and von Haefen, R.H. (2009). Distributional and Efficiency Impacts of Increased U.S. Gasoline Taxes. *The American Economic Review*, forthcoming.
- Dahl, C.A. (1979). Consumer Adjustment to a Gasoline Tax, *The Review of Economics and Statistics*, 61(3), 427-432.
- Deaton, A. and Muellbauer, J. (1980). *Economics and consumer behavior*, Cambridge University Press, reprinted 1999.
- EIA (2008a). *Annual Energy Review 2007*, Energy Information Administration, U.S. Department of Energy.
- EIA (2008b). *Monthly Energy Review*, April 2008, Energy Information Administration, U.S. Department of Energy.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.
- Hall, P. and Huang, L.-S. (2001). Nonparametric Kernel Regression Subject to Monotonicity Constraints, *The Annals of Statistics*, 29(3), 624-647.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Econometric Society Monograph Series 19, Cambridge University Press.
- Härdle, W. and Marron, J.S. (1991). Bootstrap Simultaneous Error Bars for Nonparametric Regression, *The Annals of Statistics*, 19(2), 778-796.
- Hausman, J.A. and Newey, W.K. (1995). Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss, *Econometrica*, 63(6), 1445-1476.
- Hughes, J.E., Knittel, C.R. and Sperling, D. (2008). Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand, *The Energy Journal*, 29(1), 93-114.
- Nadaraya, E.A. (1964). On Estimating Regression, *Theory of Probability and its Applications*, 9(1), 141-142.
- ORNL (2004). *2001 National Household Travel Survey. User Guide*, Oak Ridge National Laboratory. Available at <http://nhts.ornl.gov/2001/>.

- Poterba, J.N. (1991). Is the Gasoline Tax Regressive?, *NBER Working Paper*, 3578.
- Ramsey, J. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis, *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2), 350-371.
- Schmalensee, R. and Stoker, T.M. (1999). Household Gasoline Demand in the United States, *Econometrica*, 67(3), 645-662.
- Watson, G.S. (1964). Smooth Regression Analysis, *Sankhya: The Indian Journal of Statistics, Series A*. 26(4), 359-372.
- West, S.E. (2004). Distributional effects of alternative vehicle pollution control policies, *Journal of Public Economics*, 88, 735-757.
- Yatchew, A. and No, J.A. (2001). Household Gasoline Demand in Canada, *Econometrica*, 69(6), 1697-1709.